DOCUMENT RESUME

ED 306 290 TM 013 152

AUTHOR Bowman, Harry L.
TITLE Teacher Licensure Test Validation and Performance Standard Studies: A Two-State Comparison.
PUB DATE Feb 89
NOTE 14p.; Paper presented at the Annual Meeting of the Louisiana Education Research Association (New Orleans, LA, February 16-17, 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; Comparative Analysis; Content Analysis; Elementary Secondary Education; Higher Education; Knowledge Level; *Licensing Examinations (Professions); Minimum Competency Testing; Public School Teachers; *Scoring; Standards; State Programs; *State Standards; *Teacher Certification; Teacher Educators; Testing Programs; *Test Validity
IDENTIFIERS National Teacher Examinations; *Standard Setting; *Teacher Competency Testing

ABSTRACT

Comparative data and results of validation and standard-setting studies are presented for five subject-matter tests for teacher certification in two states. The five tests evaluated were available from the Educational Testing Service (ETS): one National Teacher Examination specialty area test (special education); and four tests from a consortium of state-sponsored testing programs and ETS (teaching hearing impaired students; teaching visually impaired students; health education; and marketing and distributive education). In each state, personnel from higher education reviewed test content, representatives of local school districts made judgments about job relevance, and groups of both from each specialized area estimated performance on each item for minimally qualified applicants. Panel sizes were: (1) content--State A, 6 to 11, and State B, 6 to 11; (2) job relevance--State A, 8 to 11, and State B, 9 to 13; and (3) knowledge--State A, 15 to 19, and State B, 18 to 22. In each state a Standards Committee--State A (n=14) and State B (n=17)--of lay members and educators made validity decisions and recommended minimum qualifying scores. In State A a minimum scaled score was recommended for each of the five tests, but in State B a recommendation was only made for the special education test, as it was considered that normative data for the other tests were not yet adequate. These results, especially the different approaches of the two Standards Committees, emphasize the importance of determining validity and minimum performance standards for teacher certification examinations at the state level. Six tables present study data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306290

TEACHER LICENSURE TEST VALIDATION AND PERFORMANCE STANDARD STUDIES: A TWO-STATE COMPARISON

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Harry L. Bowman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Harry L. Bowman
Executive Secretary
Mid-South Educational Research Association
College of Education
Memphis State University
Memphis, Tennessee 38152

Paper presented at the Annual Meeting of the Louisiana Education Research Association; New Orleans, LA; February 16-17, 1989

TM013152

BEST COPY AVAILABLE

Introduction

The use of subject-matter tests for the initial licensure of public school teachers is required by statute in many states. A prerequisite to the utilization of tests in this manner is a state-level study to determine their validity for the applicable subject-matter/teaching endorsement/professional service areas and the minimum scores required for applicants. Decisions regarding the validity and minimum qualifying scores of tests must be based on data derived from systematic review of the tests by qualified professional personnel from public education agencies and higher education institutions.

In 1988, independent studies were conducted through the Bureau of Educational Research and Services at Memphis State University for two states to provide data on the validity and minimum qualifying scores regarding five subject-matter tests common to both studies. The purpose of this paper is to present comparative data and results of the validation and standards-setting studies of the five tests for the two states in the Mid-South region. Each study for the respective states had two objectives: (1) to determine the validity of the tests as subject-matter instruments for endorsements of public school personnel and (2) to formulate recommendations on minimum qualifying scores for the valid tests.

Description of Candidate Tests

The five tests included in both studies are secured instruments that are available from Educational Testing Service (ETS), a private, non-profit testing organization. The Special Education Test is an NTE Sepcialty Area test that became available for administration in 1985. ETS works with a consortium of states with state-sponsored testing programs for teacher licensure to develop tests for particular specializations that are not included among the NTE Specialty Area tests. Recently developed state-sponsored testing program

(SSTP) tests available for these studies were tests for Teaching Hearing Impaired Students, Teaching Visually Impaired Students, Health Education, and Marketing and Distributive Education.

Strategy

The methodology employed in each study involved groups or panels of higher education institutional personnel and local school district professional staff with the panelists from both sources having expertise in the areas related to the respective tests. Higher education personnel reviewed the test content to make judgments about (1) the opportunity of students completing the related program to acquire the knowledge and academic skills to answer the questions correctly and (2) the congruence between the test content and the curriculum required for professional preparation. Local school district specialists made judgments about the degree to which the knowledge and academic skills covered by the test items related to competent performance as a beginning professional practitioner in the specialized area. Both groups from each specialized area estimated the performance level on each item (correct response rates of 2%, 10%, 25%, 40%, 60%, 80%, or 98%) for minimally qualified personnel who are initial endorsement applicants. A current form of each ETS test was supplied by Educational Testing Service for review by the panel members who worked independently under the supervision of study staff and ETS personnel.

A Standards Committee composed of broadly representative professional and lay personnel from within each state was convened to make decisions regarding test validity based on the content review and job relevance review data. The committee was also assigned the responsibility to recommend minimum qualifying scores for valid tests utilizing knowledge estimation results and test examinee performance data.

Participant Nomination and Selection

The nomination of potential panel members was solicited by letter that specified the number of nominees to be identified for each area of specialization. Nominations were requested from the chief academic administrators of the teacher education units in the respective states with approved preparation programs for the applicable specializations. The superintendents of representative school districts throughout each state were requested to make nominations. The instructions to the nominators specified minimum qualifications with respect to current assignment, experience, and preparation in the specialized area.

The selection of panel members for each study was performed after ascertaining that the nominees met the stated criteria for nomination. To the extent possible within each area of specialization, nominees were selected on the basis of expert qualifications, representation of gender groups, inclusion of relevant racial groups, and representation of institutions offering specialized preparation programs in the state.

Presentation of Data

In this section, the data are presented on the panel participants and the responses of the panelists in performing the assigned functions. The information from each study are reported separately by state (State A and State B).

Data on Panels

Table 1 contains the data on the number of participants for panel functions by test and state. In the study for State A, the number of panelists varied from 6 to 11 for the content review, 8 to 11 for the job relevance review, and 15 to 19 for the knowledge estimation. The study for State B

Table 1

Number of Participants for Panel Functions by Test and State

Test (Number and Name)	Panel Function: Content Review N State A	Content Review N State B	Job Relevance N State A	Job Relevance N State B	Knowledge Estimation N State A	Knowledge Estimation N State B
27 Teaching Hearing Impaired Students	7	9	8	9	15	18
28 Teaching Visually Impaired Students	7	6	9	13	16	19
35 Special Education	8	10	11	9	19	19
55 Health Education	10	11	9	12	19	22
56 Marketing and Distributive Education	6	8	9	11	15	19

included 6 to 11 panelists for the content review, 9 to 13 panelists for the job relevance review, and 18 to 22 panelists for the knowledge estimation. The larger numbers of participants for the knowledge estimation reflect the use of both the content reviewers and the job relevance reviewers to perform this function.

Content Review Panel Results

The content review of the ETS tests was conducted for both the test items and the test topics. At the test item level, judgments were made on whether or not 90% of the students completing the appropriate preparation program would have the opportunity to acquire the knowledge or academic skills to answer the question correctly. The results of the analyses are reported in Table 2.

The data indicated that over 50% of the panelists in each state viewed at least 90% of the items as content appropriate (90% to 98% for State A and 92% to 99% for State B). Based on the criterion of more than 70% of the panelists,

Table 2

Summary of Data Analyses on Content Review Ratings of Test Items by Test and State*

Test (Number and Name)	No. of Items	State	Percentage of College Personnel Who Rated Item Content as Appropriate for Curriculum					
			Over 50%	Over 60%	Over 70%	Over 80%	Over 90%	N
27 Teaching Hearing Impaired Students	120	A	90	87	87	71	37	7
		B	99	98	89	74	45	9
28 Teaching Visually Impaired Students	120	A	93	85	85	73	48	7
		B	93	93	77	77	44	6
35 Special Education	150	A	95	95	92	81	55	8
		B	94	90	79	57	32	10
55 Health Education	120	A	90	83	72	60	31	10
		B	92	87	82	73	50	11
56 Marketing and Distributive Education	120	A	98	98	94	94	78	6
		B	94	94	78	48	15	8

*Percentages may not equal 100% for a test due to omissions.

the percentages of test items judged content appropriate varied from 72% to 89% (72% to 94% for State A and 77% to 89% for State B). The percentages tended to be slightly higher for State A on the Special Education Test and the Marketing and Distributive Education Test while slightly higher percentages were observed for State B on the Teaching Hearing Impaired Students Test and Health Education Test.

Comparisons were made of the major test content topics and the emphases in the related preparation program. Allowing a difference of 5% or less, the relative emphases were rated as the "same as," "more than," or less than." The responses for each test were pooled by topic, weighted by percentage of test items on the topic, and summarized by computing a Difference in Relative

Emphasis (DRE) Index with a possible range from 0 (very close similarity) to 100 (very little similarity). The results are summarized in Table 3.

Table 3

Summary of Data Analyses on Comparisons of Curriculum Content and Test Topic Emphasis by Test and State

Test (Number and Name)	State	Difference in Relative Emphasis Index*	N
27 Teaching Hearing Impaired Students	A	29.0	7
	B	28.8	8
28 Teaching Visually Impaired Students	A	4.3	7
	B	14.4	6
35 Special Education	A	26.2	8
	B	20.6	9
55 Health Education	A	12.0	10
	B	19.1	11
56 Marketing and Distributive Education	A	31.4	6
	B	47.3	8

*Index range: 0 to 100
Low score indicates close similarity; high score indicates little similarity.

The DRE Index for the five tests varied from 4.3 to 47.3 in the two studies. The range for State A was 4.3 (Teaching Visually Impaired Students Test) to 31.4 (Marketing and Distributive Education Test). The range for state B was 14.4 (Teaching Visually Impaired Students Test) to 47.3 (Marketing and Distributive Education Test). When viewed by test, the values of the DRE Index were higher in the study for State B than in the study for State A on three tests. The DRE Index values were relatively comparable on the other two tests for both studies.

Job Relevance Panel Results

The items on the ETS tests were reviewed to judge their relevance to the responsibilities of beginning specialized professional public educators in the respective states. Responses of "Crucial" and "Important" were classified as signifying relevance while responses of "Questionable" and "Not Relevant" were defined as indicating irrelevance. The results of the job relevance ratings for the tests are presented in Table 4 for the two studies.

Table 4

Summary of Data Analyses on Job Relevance Ratings of Test Items by Test and State*

Test (Number and Name)	No. of Items	State	Percentage of Public School Personnel Who Rated Item Content as Relevant to Job: Over 50%	Over 60%	Over 70%	Over 80%	Over 90%	N
27 Teaching Hearing Impaired Students	120	A	93	93	87	76	45	8
		B	98	89	98	59	30	9
28 Teaching Visually Impaired Students	120	A	89	86	77	68	50	9
		B	96	93	86	78	66	13
35 Special Education	150	A	91	86	77	64	51	11
		B	90	86	78	64	44	9
55 Health Education	120	A	96	86	76	55	37	9
		B	87	81	72	59	48	12
56 Marketing and Distributive Education	120	A	98	93	88	80	50	9
		B	98	96	94	88	66	11

*Percentages may not equal 100% for a test due to omissions.

The ratings for the tests revealed that more than 50% of the panelists considered at least 87% of the items as job relevant (89% to 98% for State A and 87% to 98% for State B). Using a criterion of over 70%, the percentages of

items judged job relevant ranged from 72% to 94% (76% to 88% for State A and 72% to 94% for State B). Although relatively similar patterns of percentages were evident for each test in the two studies, greater differences were found among the five tests for the larger criteria primarily.

Knowledge Estimation Panel Results

The responses of panel members in performing the knowledge estimation function were analyzed to derive estimated raw score means for minimally qualified licensure applicants seeking endorsement in the specialized areas related to the tests. Using conversion factors provided by ETS, the scaled score equivalent of the raw score mean adjusted for guessing correctly was calculated for each test. The results for the knowledge estimation function are reported in Table 5.

Table 5

Summary of Analyses on Estimated Scores for Minimally Qualified Certification Candidates by Test and State

Test (Number and Name)		Items Total/Scored	State	Mean Raw Score	Mean Scaled Score	N
27	Teaching Hearing Impaired Students	120/120	A	69.473	557.014	15
			B	66.175	531.364	18
28	Teaching Visually Impaired Students	120/120	A	65.730	527.899	16
			B	72.359	579.463	19
35	Special Education	150/148	A	76.730	556.181	19
			B	73.525	543.612	19
55	Health Education	120/119	A	70.325	568.235	19
			B	60.594	491.914	22
56	Marketing and Distributive Education	120/120	A	71.124	569.858	15
			B	69.459	556.902	19

The scaled scores for the ETS subject-matter tests can vary from a low of 250 to a high of 990, a 740-point difference between the lowest and highest scores possible. Scores cannot be compared directly among the subject-matter tests for two reasons. First, the tests are normed independently on different groups of examinees representing different populations. Second, the standard errors of measurement, an index of precision of scores on each test, vary considerably among the tests. An inspection of the scaled score equivalents for the raw score mean on each test revealed that the degree of similarity between the states ranged from high to relatively low.

Standards Committee Actions

The Standards Committee for each study was composed of lay members and professional personnel from higher education institutions and public education agencies. The committee consisted of 14 members in the study for State A and 17 members in the study for State B. Each committee elected one of its members to serve as chair and followed the same steps in performing the assigned tasks.

Test Validity Decisions

Each Standards Committee reviewed the analyses of the data collected in the performance of the content review and job relevance review functions for the five ETS tests. These data focused on the congruence between (1) the test content and the related professional programs in the state and (2) the test content and the knowledge and academic skills required for success as a beginning professional practitioner in the state. These data were presented and explained to the committee by the study staff. The committee considered both the content review and job relevance review results in making decisions on the validity of each test. Based on a thorough review and discussion of the data, each committee decided that each test was valid to use as an initial

licensure test in their respective states. The rationale for these decisions by each committee was that the contents of the tests had relatively high congruence with the related professional preparation programs and addressed knowledge and academic skills needed for competent performance as beginning professional practitioners in the specialized areas.

Recommended Minimum Qualifying Scores

After making affirmative decisions on test validity, each Standards Committee received the results of the knowledge estimation function performed by the panelists assigned to the respective tests. These data were essential to determine the minimum qualifying scores that should be required for initial licensure applicants seeking endorsements in the specialized areas related to the tests. Additional information presented to the committee was the examinee performance data available from ETS. The number of examinees for the Special Education Test was sufficiently large to provide normative data, including the standard error of measurement. The relatively small number of examinees who had taken the other tests - from 5 to 66 - did not permit the derivation of normative information, thereby limiting the utility of these data in providing guidance to the committee. Standard errors of measurement available for similar ETS tests - generally varying from about 15 to 28 in scaled score units - were provided for guidance in the committee's deliberations.

After a comprehensive review and discussion of the data collected from panelists and the other available information, each committee made decisions regarding the recommended minimum qualifying scores for each test to be used in their respective states. The recommendations are reported in Table 6. In the study for State A, minimum scaled scores were recommended for each of the five tests. In the study for State B, a minimum scaled score was recommended for only the Special Education Test that had a normative database. The committee

Table 6

Recommended Minimum Qualifying Scores on the ETS Tests Studied for Initial Licensure by State

	Required Score State A	Required Score State B
Teaching Hearing Impaired	500	no minimum
Teaching Visually Impaired	480	no minimum
Special Education	510	490
Health Education	520	no minimum
Marketing and Distributive Education	520	no minimum

for the latter study recommended that all applicants for initial licensure with endorsements in the areas corresponding to the four valid tests without a normative database should be required to submit scores on the appropriate tests without any specific minimum qualifying scores being established.

Other Recommendations

Additional recommendations were made by each Standards Committee in response to specific concerns that arose during the deliberations. The committee for the study in State A recommended that the State Board of Education monitor the results of using the five ETS tests during the first two years of implementation and review the appropriateness of the score requirements for initial licensure based on the four tests without normative data. The committee for the study in State B made two recommendations that included attention to concern about potential adverse impact on legally protected minorities: (1) to monitor the performance of examinees on the

Special Education Test to determine the impact on special groups of examinees and (2) to collect data on examinee performance for the four required tests without established minimum scores, these data to include racial group identity, until sufficient data are available to set minimum qualifying scores.

Conclusions

The results of the two studies, especially the differences in decisions on minimum qualifying scores, demonstrate the importance of determining validity and minimum performance standards for teacher licensure tests at the state level. Since teacher licensure requirements are established and implemented by state licensing agencies, these agencies need to be assured that the tests and minimum scores used for licensure are appropriate if challenges of the requirements are made. These studies also indicate that the ETS subject-matter tests may have broadly geographic validity as teacher licensure tests for the applicable specialized areas. The establishment of minimum performance standards, however, may be affected by the magnitude of the examinee performance database available for valid tests and the nature of the concerns by each Standards Committee regarding the impact of establishing minimum performance standards.